

**UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA ANIMAL**



STUDIES ON THE INFLUENCE OF A TO I RNA EDITING ON PROTEIN EVOLUTION

VERSÃO PÚBLICA

Diogo de Abreu Marques Ribeiro

Dissertação

Mestrado em Bioinformática e Biologia Computacional
Especialização em Bioinformática

2012

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA ANIMAL



STUDIES ON THE INFLUENCE OF A TO I RNA EDITING ON PROTEIN EVOLUTION

VERSÃO PÚBLICA

Diogo de Abreu Marques Ribeiro

Dissertação de mestrado orientada por
Doutor Alekos Athanasiadis (Instituto Gulbenkian de Ciência)
e Prof. Doutor Octávio Paulo (Faculdade de Ciências da Universidade de Lisboa)

Mestrado em Bioinformática e Biologia Computacional
Especialização em Bioinformática

2012



FUNDAÇÃO CALOUSTE GULBENKIAN
Instituto Gulbenkian de Ciência

STUDIES ON THE INFLUENCE OF A TO I RNA EDITING ON PROTEIN EVOLUTION

Diogo de Abreu Marques Ribeiro

MASTER THESIS

2012

This thesis was fully performed at Protein-Nucleic Acids Interactions Laboratory, Instituto Gulbenkian de Ciência under the direct supervision of Dr. Alekos Athanasiadis.

Prof. Dr. Octávio Paulo was the internal supervisor in the scope of the *Master in Bioinformatics and Computational Biology* of the Faculty of Sciences of the University of Lisbon.

Resumo

A edição de ARN é um processo que altera as propriedades de moléculas de ARN – incluindo ARN transportadores, ARN mensageiros e micro-ARNs. A edição de ARN ocorre em eucariotas e abrange vários tipos, como inserção ou remoção de nucleótidos (uracilos) e dois tipos de desaminação de nucleótidos: de citosina para uracilo e de adenosina para inosina. A edição de ARN de adenosina para inosina (edição de ARN A-para-I) é o tipo de edição de ARN mais comum em metazoários e ocorre em transcritos de ARN possuindo estruturas secundárias, onde adenosinas são convertidas a inosinas por uma desaminação catalizada por enzimas da família adenosina desaminase específica do ARN (ADAR). Como a inosina forma pares de bases com citosinas, esta é interpretada pela maquinaria celular como sendo guanosina, mudando assim o conteúdo informativo de uma molécula de ARN mensageiro, e consequentemente a funcionalidade proteica. Presume-se que este processo ocorra em mais de 85% dos ARN pré-mensageiros humanos, conhecendo-se pelo menos 50 proteínas modificadas por este mecanismo. Para além disso, a edição de ARN A-para-I é fundamental em várias espécies pois a não-edição de alguns transcritos está correlacionada com várias doenças e deficiências – podendo levar à morte prematura de camundongos ou causar deficiências comportamentais em *Drosophila melanogaster* e *Caenorhabditis elegans*. Em humanos este fenómeno está associado com cancro, esclerose lateral amiotrófica, diabetes e lúpus, entre outros. Embora se conheça relativamente bem o mecanismo de acção da edição de ARN A-para-I, como também a sua prevalência e consequências biológicas, o efeito que este processo tem na evolução dos seus substratos permanece desconhecido.

Sabe-se que os mecanismos epigenéticos, por estes gerarem uma grande diversidade transcriptómica, são factores essenciais na sustentabilidade da imensa complexidade proteica dos organismos superiores. A edição de ARN mostra ter um grande potencial para fornecer parte desta diversidade e foi sugerido que este mecanismo possa aumentar a evolvabilidade de populações e – por ser abundante no sistema nervoso – que seja um factor relevante por exemplo na evolução das capacidades cognitivas humanas. Para elucidar estas questões de uma maneira extensiva e controlada – dado o número de estudos evolutivos deste mecanismo ser reduzido, e por a evolução ser um processo largamente imprevisível – neste trabalho estudamos a evolução de genes codificantes de proteína sobre edição de ARN, usando métodos *in silico*. O nosso propósito é descobrir como a edição de ARN altera o curso de evolução dos seus alvos e por quais mecanismos isto ocorre, através de uma simulação realista da evolução de populações digitais na

presença de edição de ARN, tendo em conta o processamento de informação genética *in vivo* (transcrição e tradução) e selecção natural em populações. Para isso utilizamos um algoritmo genético modificado para albergar um passo entre a transcrição e tradução de ARN onde este pode sofrer edição de adenosina para inosina e consequentemente alterar a similaridade entre a proteína produzida e uma hipotética proteína-alvo. O processo de edição de ARN *in vivo* varia largamente na sua eficiência – a presença de um evento de edição pode variar entre 1% a 100% de todos os transcritos produzidos – e também na sua especificidade, onde metade de todas as adenosinas num transcrito podem ser editadas, ou apenas uma ou duas adenosinas específicas serem editadas. Para estudar o efeito da edição de ARN nestas variadas condições, inserimos vários níveis de complexidade nas nossas simulações de forma a retratar evolução em diferentes modos de edição e diferentes condições ambientais (e.g. evolução em direcção a uma proteína-alvo variável ou evolução num ambiente fixo).

Com estas simulações descobrimos que a evolução populacional, relativamente à produção de uma proteína-alvo fixa, é abrandada quando estas populações suportam elevados níveis de edição de ARN, comparativamente a populações evoluindo sem a presença de mecanismos de edição. Este abrandamento é devido a constrangimentos evolutivos das sequências codificantes envolvidas na manutenção de eventos de edição. Mesmo que um evento de edição realize uma alteração benéfica para o organismo, a manutenção de um evento de edição requer a presença de duas sequências complementares no transcrito – mimetizando os requerimentos estruturais *in vivo* presentes na ligação das proteínas ADAR aos alvos. Assim, mutações que modifiquem sequencias que sustentem um evento de edição dificilmente poderão conferir benefícios para o organismo, caso essas mesmas mutações também anulem um evento de edição benéfico para o organismo.

Contudo, quando os níveis de edição de ARN são moderados ou baixos, as populações não evoluem tão dependentes de eventos de edição benéficos, e a presença de edição de ARN é até capaz de aumentar ligeiramente a taxa de evolução (evolvabilidade) das populações quando comparada a um controlo evoluindo sem edição de ARN. O impacto da edição de ARN nas taxas de evolução advém da alteração do balanço entre três forças evolutivas que actuam na presença deste mecanismo: constrangimentos na evolução das sequencias que sustentam eventos de edição benéficos para o organismo; abrandamento da evolução causado pela remoção sistemática de eventos de edição de ARN prejudiciais; aumento da exploração do espaço genético através da edição de ARN. Esta última força evolutiva tem a capacidade de aumentar a capacidade evolutiva das populações através de um processo conhecido como o efeito Baldwin. Esta teoria conjectura que a presença de um novo fenótipo (aprendido ou plástico) numa população pode alterar a pressão selectiva

dessa população, beneficiando indivíduos que mais facilmente adquiram essa característica. Subsequentemente, ainda que esse fenótipo não seja directamente determinado geneticamente, a modificação da pressão selectiva vai provocar assimilação genética desse fenótipo ou de características que viabilizem uma mais fácil aquisição desse fenótipo. Nas nossas simulações observamos esse efeito; populações sob edição de ARN evoluem mais rápido porque, em codões que só estão a uma modificação (a um nucleótido) de codificar o aminoácido-alvo, a edição de ARN pode imediatamente converter este codão que de outra maneira seria “inapto” num codão certo que contribui para a *fitness* do organismo. Assim a pressão selectiva na população é alterada, onde organismos com codões mais perto de codificar o aminoácido certo são preferencialmente seleccionados. De seguida, verificamos uma rápida assimilação genética de eventos de edição, pois a codificação genética é mais vantajosa por diminuir a dependência dos organismos em eventos de edição, visto que a modificação de adenosina para inosina é apenas dada por uma chance, não ocorrendo em todos os transcritos.

Neste trabalho abordamos também o efeito da edição de ARN em populações em equilíbrio num ambiente estável. Sob estas condições a edição de ARN mostra aumentar a variabilidade de genótipos da população, ao mesmo tempo que mantêm o mesmo valor médio de similaridade com a proteína-alvo que populações sem edição. Esta maior variabilidade genotípica mostra ser importante em ambientes mutáveis, onde populações sob edição de ARN mostram adaptar-se mais rapidamente a alterações na proteína-alvo do que populações controlo. Simulações usando uma função alternativa para o cálculo da *fitness* – baseada numa matriz de substituição de aminoácidos – provoca estagnação evolutiva de populações que atingiram um pico sub-óptimo de *fitness*, contudo, populações com edição de ARN demonstram uma maior capacidade de escapar estagnação evolutiva e alcançar maior *fitness*. A edição de ARN mostra assim ser capaz explorar espaço genotípico inacessível pela mutação e recombinação conseguindo alcançar fenótipos que de outra maneira eram inacessíveis.

Em suma, embora este trabalho não determine a extensão do impacto do mecanismo de edição de ARN A-para-I em populações *in vivo*, estas simulações dão-nos pistas de como este mecanismo poderá afectar evolução de genes codificantes, revelando a potencial contribuição deste mecanismo epigenético para a adaptação evolutiva, não só aumentando a evolvibilidade de uma população mas também explorando o espaço evolutivo de uma maneira dissimilar.

Palavras-chave: edição de ARN; adaptação; evolvibilidade; Efeito Baldwin; algoritmo genético; evolução *in silico*

Abstract

Adenosine-to-inosine RNA editing is a pervasive epigenetic mechanism that increases transcriptome diversity in metazoans. The phenomenon occurs posttranscriptionally targeting intramolecular stem-loops, where adenosine (A) is converted to inosine (I) by adenosine deaminases that act on RNA (ADARs). As inosines are interpreted by the cellular machinery as guanosines, the information content of the mRNAs is modified, possibly leading to altered protein function. Although the prevalence, molecular mechanism and physiological function of A-to-I RNA editing is understood to some extent, its evolutionary significance still remains a mystery. Here, we simulate evolution of digital populations under RNA editing in several relevant conditions (e.g. different specificity and efficiency) in order to study the dynamics and evolution of the RNA editing targets in an extensive and controlled way. We aim to discover the specific mechanisms through which RNA editing affects protein-coding gene's evolution by simulating the *in vivo* information processing and natural evolution. We found that high levels of RNA editing, even when improving population fitness, may slowdown population evolution, mainly caused by the requisite of having to sustain complementary sequences within the gene for RNA editing to occur. However, low levels of RNA editing increase the evolvability of the system because RNA editing augments the sequence space exploration leading to enhanced genetic assimilation of improved genotypes, hence representing the Baldwin Effect hypothesis. Additionally, we address the impact of RNA editing in evolution under different scenarios: changing environment, accounting for amino acid similarities, and observing the effects of RNA editing on populations at equilibrium. Overall, this work allows us to posit further value of this epigenetic source of phenotypic variation as a relevant contributor for adaptive evolution. This work also provides us clues on how protein-coding genes evolve under RNA editing and subsequent studies should aim to establish the extent to which RNA editing influences metazoan evolution.

Keywords: RNA editing; adaptation; evolvability; Baldwin effect; genetic algorithm; *in silico* evolution